Recipe for Success:

Data-Driven Strategies for New & Prospective Restaurant Owners in San Diego

By Tia Page

Overview

Introduction	Defining the problem & importance of topic
Data Collection	Resources and techniques used to obtain the data needed for analysis
Analysis of Data	Valuable findings and insights gathered through data explorations
Conclusion	Major takeaways from analysis that can help aspiring restaurant owners

Introduction

Problem: San Diego restaurant closures

Causes: Highly competitive area for restaurants, overly saturated market, high costs of living

Solution: High revenues earned as a result of attracting customers and establishing popularity

Project Goals: Identify the key factors of successful restaurants in San Diego, recommend optimal locations to establish business, and find ways to enhance customer experience

Data Collection

- **Data Collected:** Restaurant reviews, Demographic information, geographic details, restaurant characteristics
- **Collection methods:** Webscraping, Data Extraction
- **Tools:** BeautifulSoup, Octospare data extractor
- **Websites:** Google Reviews, Point2Homes.com, Bklyndesigns.com, usa.com
- **Final dataset:** sd_reviews

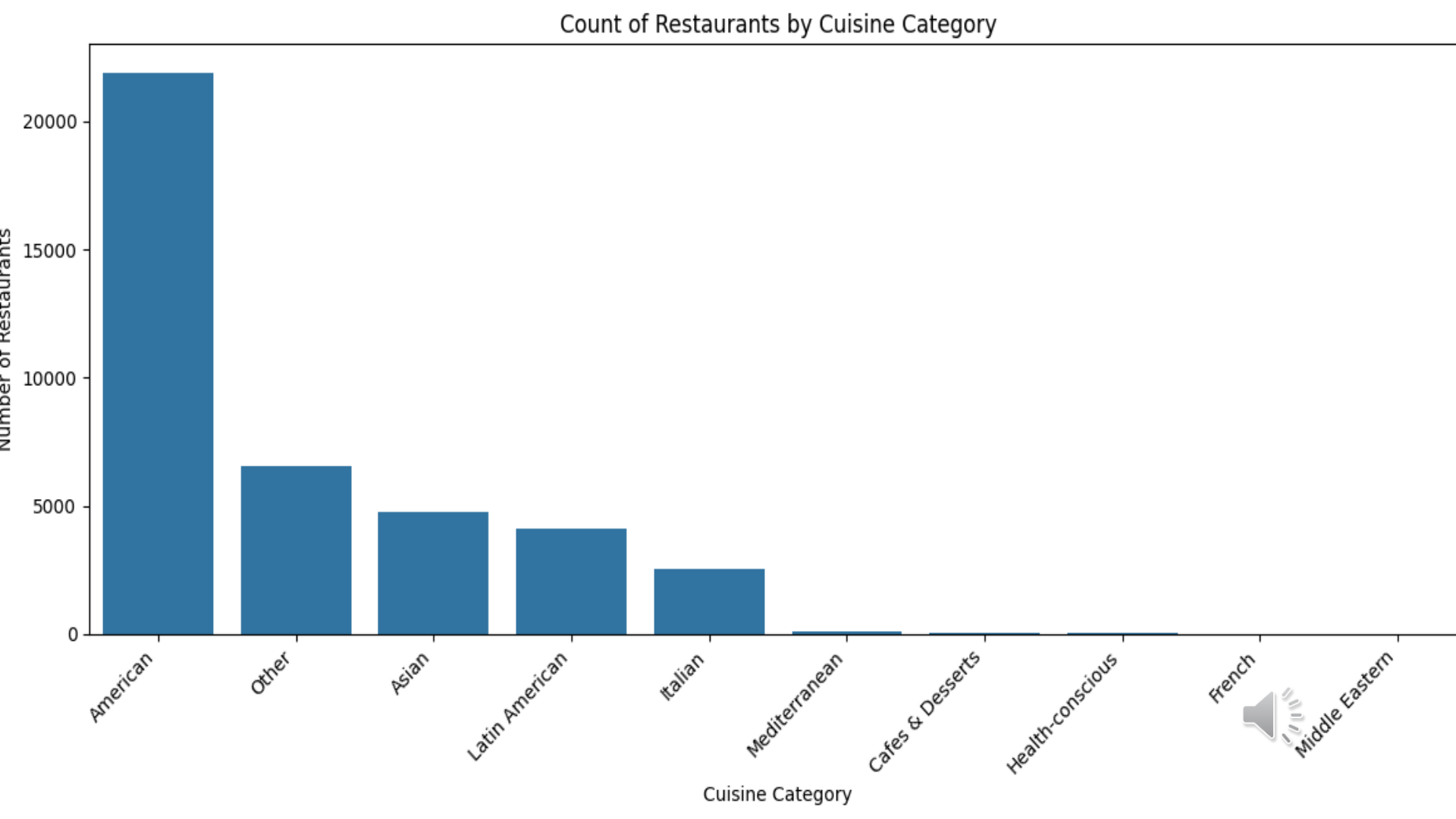
Data Cleaning & Preparation

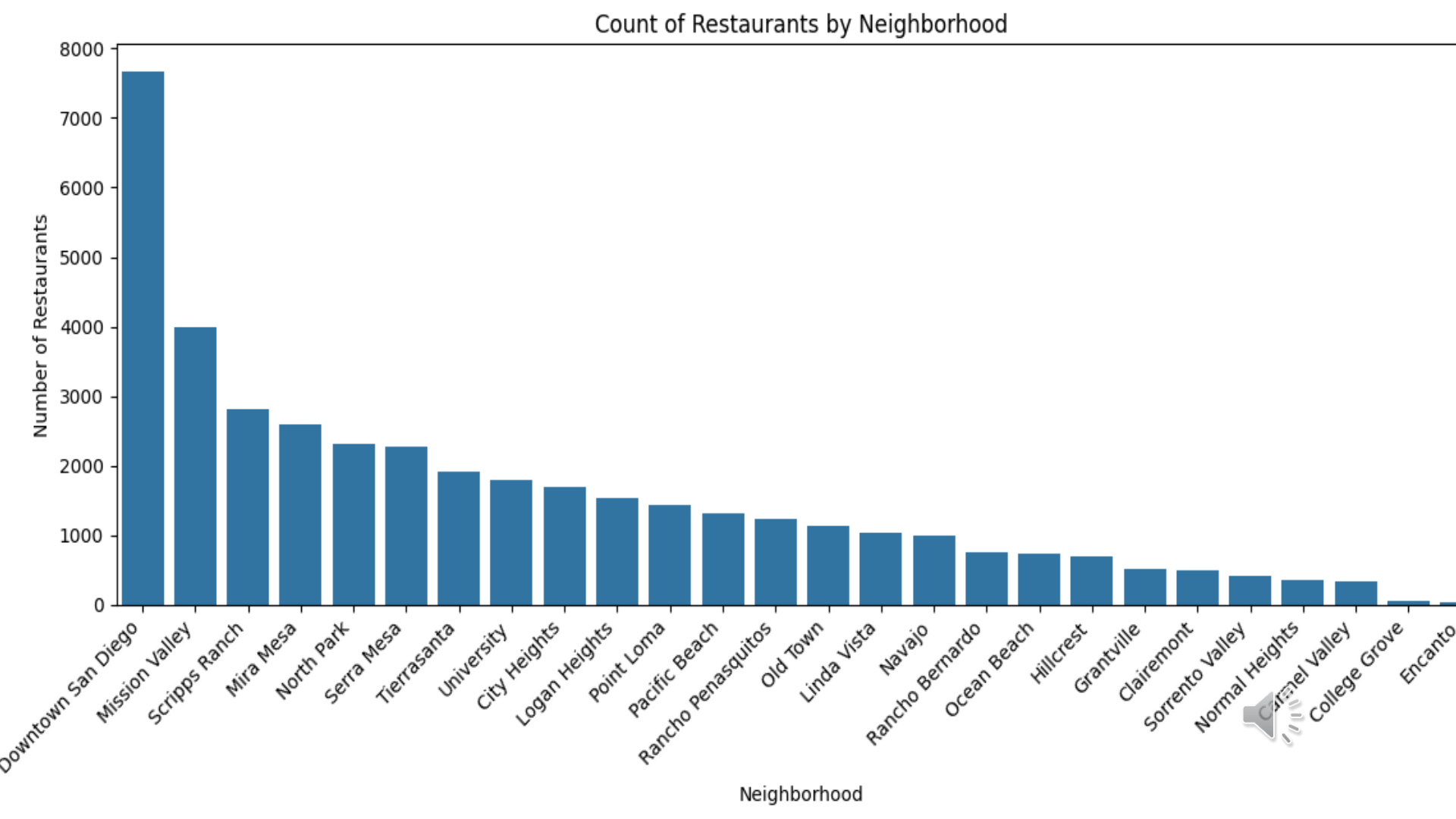
- **Removed unnecessary columns:** Review likes, review urls, review response
- **Excluded rows with missing values**
  - Some data failed to extract
- **Incorrect/missing cuisine categories replaced** using keywords in names
- **Merged** separate datasets obtained via data extraction and web scraping

EDA: Overview of Data Set

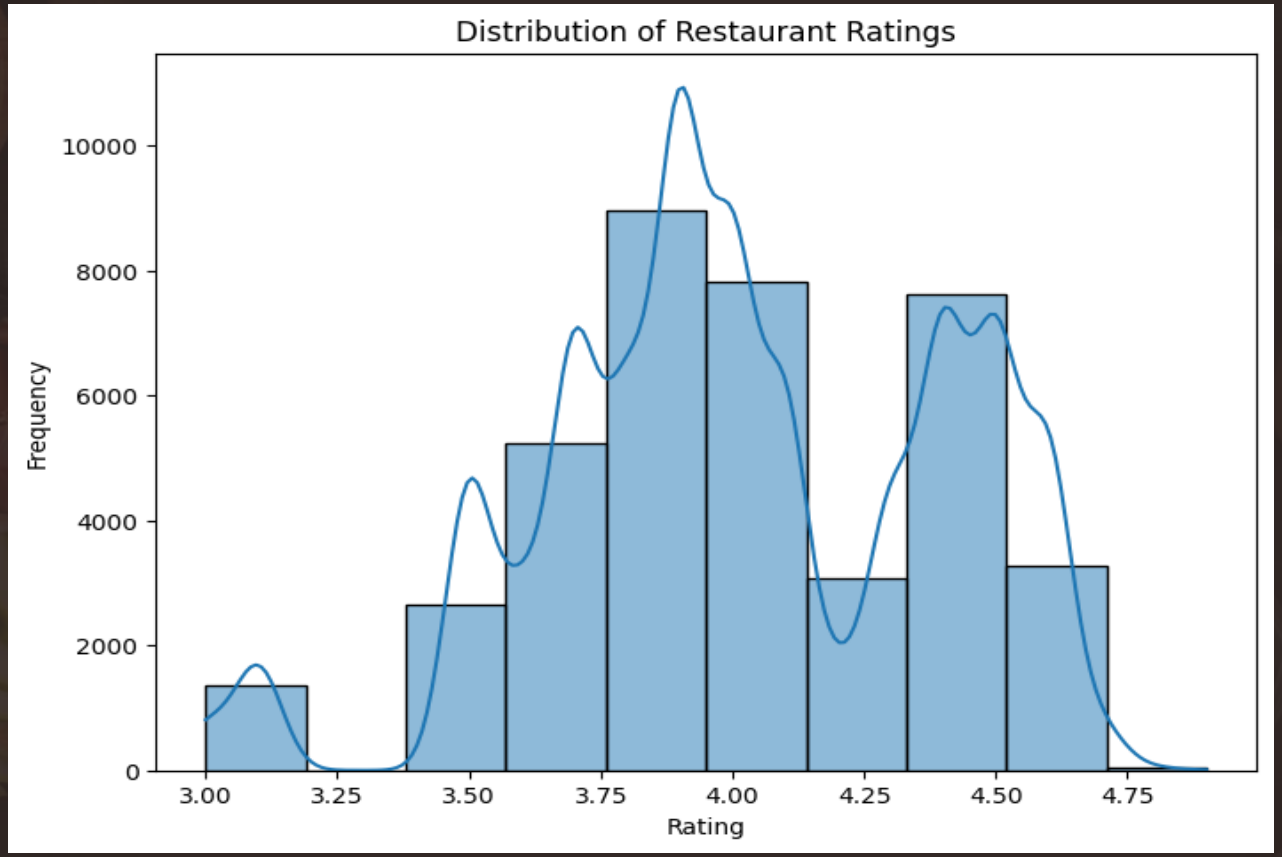
- 42,768 restaurants (rows)
- 16 variables measures (columns)
- Mixture of categorical & numeric data

```
<class 'pandas.core.frame.DataFrame'>
Index: 42768 entries, 501 to 58042
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   restaurant_name  42768 non-null  object
 1   rating           42768 non-null  float64
 2   review_count     42768 non-null  object
 3   lowest_price     42768 non-null  float64
 4   highest_price    42768 non-null  float64
 5   cuisine_type     42768 non-null  object
 6   zip_code         42768 non-null  int64
 7   neighborhood     42768 non-null  object
 8   land_area        42768 non-null  object
 9   population       42768 non-null  float64
 10  num_households   42768 non-null  float64
 11  median_income    42768 non-null  float64
 12  average_income   42768 non-null  float64
 13  review_rating    40154 non-null  object
 14  review           40020 non-null  object
 15  review_likes     9642 non-null   float64
dtypes: float64(8), int64(1), object(7)
```


Count of Restaurants by Cuisine Category
Number of Restaurants
20000
15000
10000
5000
0
American
Other
Asian
Latin American
Italian
Mediterranean
Cafes & Desserts
Health-conscious
French
Middle Eastern
Cuisine Category


Count of Restaurants by Neighborhood
Number of Restaurants
8000
7000
6000
5000
4000
3000
2000
1000
0
Downtown San Diego
Mission Valley
Scripps Ranch
Mira Mesa
North Park
Serra Mesa
Tierrasanta
University
City Heights
Logan Heights
Point Loma
Pacific Beach
Rancho Penasquitos
Old Town
Linda Vista
Navajo
Rancho Bernardo
Ocean Beach
Hillcrest
Grantville
Clairemont
Sorrento Valley
Normal Heights
College Grove
Encanto
Neighborhood

Analysis of Data



Addressing Goals

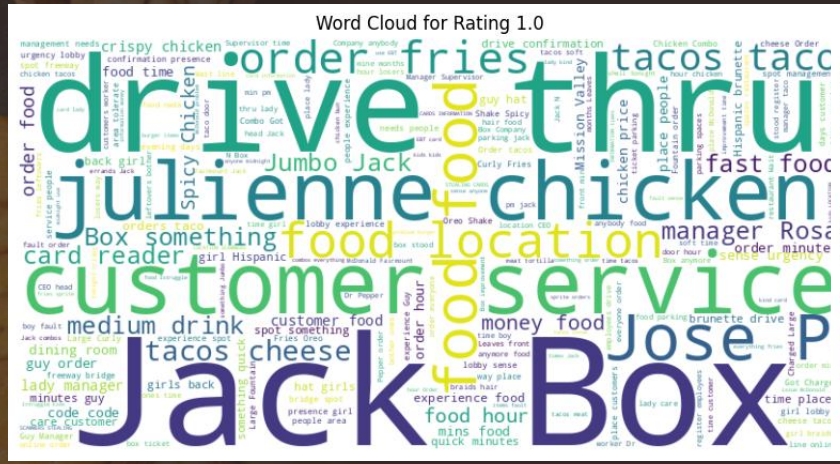
#1. Identifying key factors of successful restaurants using Random Forest Classifier Model

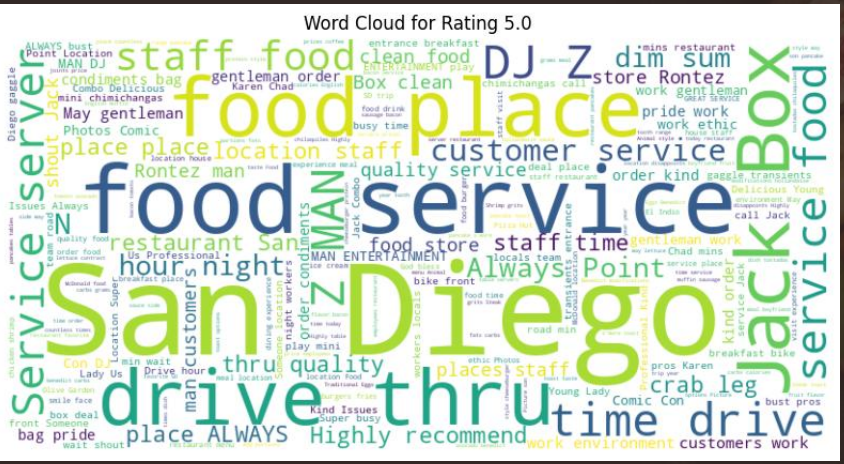
Model Accuracy: 1.0

Feature Importance:

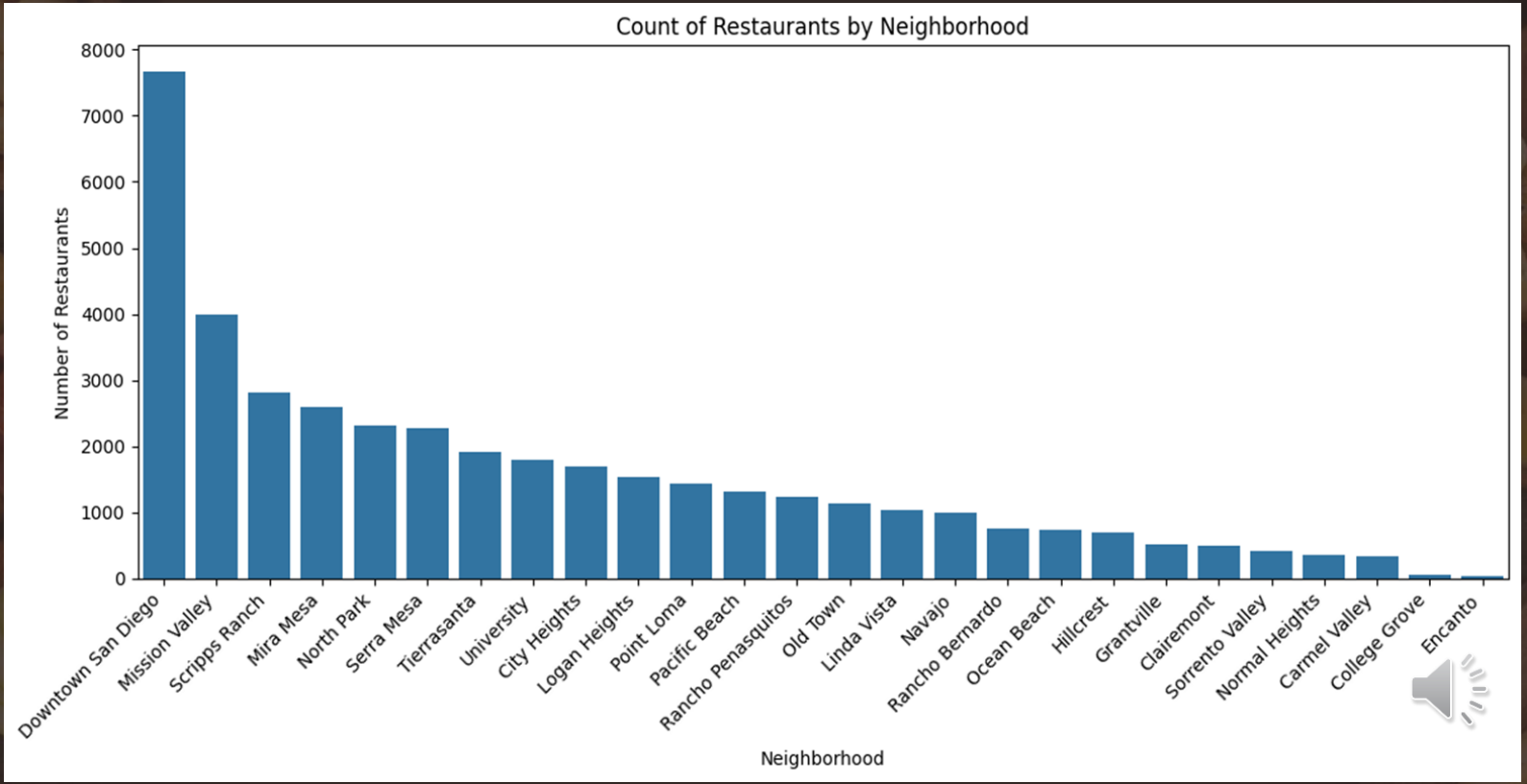
	Feature	Importance
0	rating	0.861155
1	review_count	0.105908
2	cuisine_category_Asian	0.016447
6	cuisine_category_Italian	0.008091
10	cuisine_category_Other	0.004074
7	cuisine_category_Latin American	0.003255
8	cuisine_category_Mediterranean	0.000499
3	cuisine_category_Cafes & Desserts	0.000429
5	cuisine_category_Health-conscious	0.000109
9	cuisine_category_Middle Eastern	0.000018
4	cuisine_category_French	0.000015

#2. Enhancing Customer Experience through TextBlob Sentiment Analysis Model

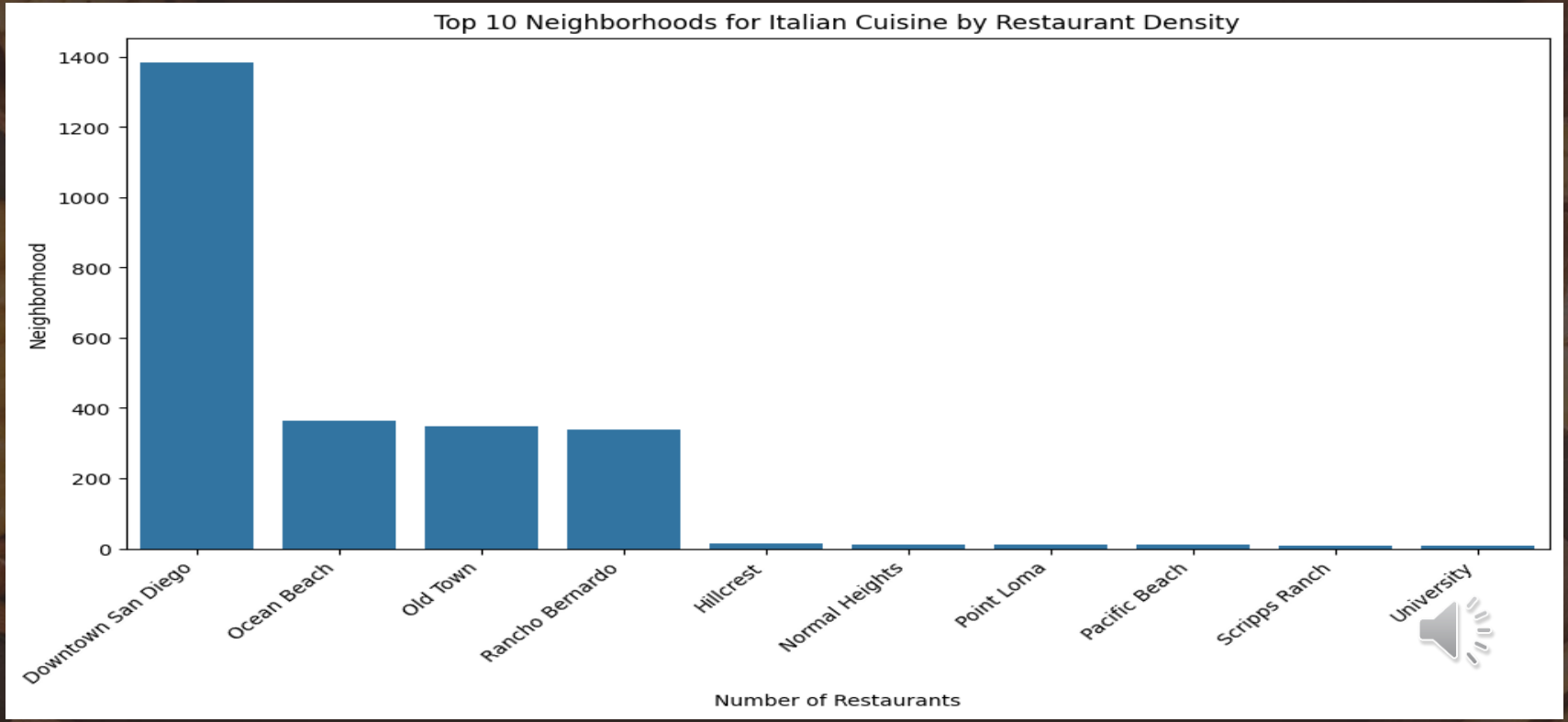




#3. Choosing an optimal restaurant location



#3. Choosing an optimal restaurant location (continued): Italian restaurant scenario



Conclusion: Main Takeaways

- Avoid competition by establishing restaurant in area with lowest density of restaurants based on specific food categories
- Enhance customer experience by first ensuring that customer service is of highest quality, then improving on food quality
- Stand out from the crowd by paying attention to the least popular food categories presented in data visualization; this also shows areas with less competition